4. BAYES METHODS

Basic concepts

Theorem 4.1.1 (Bayes' theorem)

$$f(y|x) = \frac{f(x|y)f(y)}{\int f(x|y)f(y)dy}.$$

Given a **prior density** $f(w)$ and a **model** $f(x|w)$, the **posterior density** after observing data $x_1 \ldots, x_n$ is

$$f(w|x_1, \ldots, x_n) = \frac{f(x_1|w \cdots f(x_n|w)f(w)}{f(x_1, \ldots, x_n)} \propto L_n(w)f(w)$$

Theorem 4.1.2 (Kolmogorov)
The posterior density depends on $\mathbf{x}$ only through $T(\mathbf{x})$ iff T is sufficient.

Theorem 4.1.3
Let $f(x|w)$ be a k-parameter exponential family. Then a conjugate family of distributions is the $(k+1)$-parameter exponential family obtained by treating n and $\sum_i T_j(x_i)$, $j = 1, \ldots, k$ as parameters, and $A(w), c_j(w)$, $j = 1, \ldots, k$ as variables.

Theorem 4.1.4 If the parameter set is an open subset of $\mathbf{R}^k$, then the conjugate family from theorem 4.1.3 has linear posterior expectation, in the sense that $\mathbf{E}(\mathbf{E}(X|W)|X = x) = ax + b$.

Theorem 4.1.5
(a) Let $f(\cdot|w)$ be a natural exponential family on a discrete sample space, where $\Omega = (-\infty, \bar{w})$, $\bar{w} < \infty$. Assume the prior is non-trivial and $\mathbf{E}(A'(w)|X = x) = ax + b$, $a > 0$. Then the prior has density proportional to $\exp((bw - (1-a)A(w))/a)$.
(b) If $f(\cdot|w)$ is a natural exponential family on a continuous sample space, containing an open interval, then linear posterior expectation implies exponential conjugate prior as in (a).

4.2 Estimation and decision theory

Definition 4.2.1 A **Bayes rule** against a loss function $L(w,d)$ minimizes the **Bayes risk** $\mathbf{E}L(W,d)$, where the expectation is taken with respect to the current distribution of the parameter.

Definition 4.2.2 An estimator d is **minimax** if

$$\sup_w \mathbf{E}(L(W,d)|W = w) = \inf_d \sup_w \mathbf{E}(L(W,d)|W = w)$$

Theorem 4.2.1 (Minimax theorem) If d is a Bayes rule such that $\mathbf{E}(L(W,d)|W = w)$ is constant as a function of w, then d is minimax.

Theorem 4.2.3 (Rao-Blackwell)
Let T be a sufficient statistic for w, and assume that $L(w,d)$ is strictly convex in d. Then if δ satisfies $\mathbf{E}L(w,\delta) < \infty$, the estimator $\eta = \mathbf{E}(\delta|T = t, W = w) \equiv \mathbf{E}(\delta|T = t)$ satisfies $\mathbf{E}L(W,\eta)|W = w)$ $< \mathbf{E}(L(W,\delta)|W = w)$ unless $\delta = \eta$ a.s.

Definition 4.2.3 An estimator d is **inadmissible** if there is an estimator δ such that $\mathbf{E}(L(W,\delta)|W = w) \le \mathbf{E}(L(W,d)|W = w)$ for all w, with strict inequality for at least one w. Then $\delta$is said to **dominate** d. An estimator which is not inadmissible is **admissible**.

Theorem 4.2.4
A unique Bayes estimator is admissible.

Theorem 4.2.5
An unbiased Bayes estimator for squared error loss is exact (i.e., unbiased with variance 0).